Introduction to Artificial Intelligence, Fall & Winter 2022
College of Computer Science, Zhejiang University
Problem Set 1.3: Reinforcement Learning (I)

丁尧相

2022 年 10 月 14 日

Problem 1. 请给出一个可以应用强化学习的实际问题的例子，并建模其中的 MDP (Markov Decision Process)：描述出 state space, action space, transition function, reward function 各是什么。

Problem 2. 在图 1所示 MDP 中，表格代表了 9 个 state，单元格内的数值代表了到达这一状态能够得到的 reward，假定执行动作状态转移是确定的，MDP 中的 $\gamma = 0.9$。请分别画出使用 value iteration 和 policy iteration 前 5 轮每个 state 对应的值函数以及策略的变化情况。
提示：每一轮同样画出表格，在对应的单元格内填上值函数和当前最优策略即可。初始值函数可以全部设置为 0，初始策略可以自己指定。

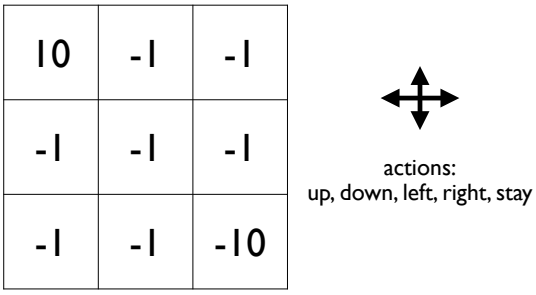

图 1: Problem 2 的 MDP。

Problem 3. (***) SARSA 和 Q-learning 都使用了 ϵ-greedy 策略进行探索，其中如何对 ϵ 进行设置是一个值得探讨的问题。请问下面几种方式是否可行？请论述你的理解。

- 在训练过程中维持一个固定的数值，如固定 $\epsilon = 0.1$。
- 在训练过程中令 ϵ 逐渐减小，但最终并不会到达 $\epsilon = 0$，而是到达一个最小数值，如 $\epsilon = 0.01$。
- 在训练过程中令 ϵ 逐渐减小，到达 0 之后再训练一些轮数，直到收敛。